

А. С. МАЗМАНИШВИЛИ, д-р физ.-мат. наук, профессор,
А. Ю. СИДОРЕНКО, канд. техн. наук, доцент каф. САиУ НТУ «ХПИ»

КОМПЕНСАЦИЯ КВАНТИЛЕЙ НАБЛЮДАЕМЫХ ДАННЫХ РАСПРЕДЕЛЕНИЯ СТЬЮДЕНТА ПРИ НАЛИЧИИ КОРРЕЛЯЦИИ

В статті розглянуто типові задачі математичної статистики, що відносяться до корельованих послідовностей (часовим відлікам) з заданим фіксованим рівнем статистичного зв'язку між ними. Одержані аналітичні вирази, що описують розподіл Стюдента у випадку корельованих відліків. Представлені таблиці квантилів розподілу Стюдента при наявності кореляції.

Рассмотрены типичные задачи математической статистики, отнесенные к коррелированным последовательностям (временным отсчетам) с заданным фиксированным уровнем статистической связи между ними. Получены аналитические выражения, описывающие распределение Стюдента для случая коррелированных отсчетов. Представлены таблицы квантилей распределения Стюдента при наличии корреляции.

The typical problems of mathematical statistics, related to correlated sequences (to time counts) with a given fixed level of statistical correlation between them, are considered. The analytic expressions featuring distributions Student's ones in the case of correlated counts are obtained. The distributions Student's quantil's in the case of correlated counts are introduction.

Введение. В теории статистического оценивания и принятия решений обычно предполагается, что элементы наблюдаемых данных (отсчетов) x_0, x_1, \dots, x_n независимы. Сюда относятся распределения χ^2 , Стюдента Фишера [1]. Между тем, возможна ситуация, когда наблюдаемые данные или данные, предназначенные для статистической обработки, в той или иной степени зависимы, то есть присутствует корреляция. Также возможно, что информация о коррелированности данных отсутствует или неизвестна к моменту обработки результатов эксперимента. Тем самым возникает проблема изучения степени влияния корреляции полученных данных на распределения рассматриваемых в задачах математической статистики.

Целью данной работы является изучение влияния корреляции на квантили, известного в математической статистике, распределения Стюдента.

Основная часть. Пусть используемые наборы отсчетов $\{x_0, x_1, \dots, x_N\}$ вместе с нулевым отсчетом x_0 случайной величины X – совокупность сечений объемом N , взятых с постоянным временным шагом из реализации нормального марковского процесса Орнштейна-Уленбека (ОУ-процесса) $x(t)$ с интенсивностью σ и коэффициентом корреляции q , а также известна последовательность $\{y_0, y_1, \dots, y_N\}$ случайной величины Y , которая является

нормальной, независимой от совокупности $\{x_0, x_1, \dots, x_N\}$ и также имеет интенсивность σ .

В данной работе, в выше указанных предположениях, использован дискретный квадратичный функционал, имеющий вид:

$$J = \frac{1}{N} \sum_{k=1}^N x_k^2. \quad (1)$$

Данный функционал имеет равновесную производящую функцию

$$Q_J(\lambda; q, n) \equiv \langle \exp(-\lambda J) \rangle_J \quad (2)$$

следующего вида [5]

$$Q_J(\lambda; q, n) = \left(\frac{(1-q^2)R}{(a_+ - q^2)^2 a_+^{N-1} - (a_- - q^2)^2 a_-^{N-1}} \right)^{1/2}, \quad (3)$$

где $a_{\pm} = (1 + q^2 + \lambda \sigma^2 (1 - q^2) / N \pm R) / 2$, $R = [(1 - q^2) + \lambda \sigma^2 / n]^2 - 4q^2$;
 q – коэффициент корреляции соседних отсчетов ОУ-процесса.

На основе (1)–(3) рассмотрена случайная величина – стандартизованное отношение Стьюдента

$$T = Y / \sqrt{J} = y_0 \left(\frac{1}{N} \sum_{k=1}^N x_k^2 \right)^{-1/2}. \quad (4)$$

Для случайной величины T статистический анализ удобно провести в терминах её характеристической функции $\Psi_T(\xi)$, имеющей вид

$$\Psi_T(\xi) = \left\langle \exp(i\xi Y / \sqrt{J}) \right\rangle_{Y,J}, \quad (5)$$

здесь индексами указаны случайные величины, по которым необходимо произвести статистическое усреднение.

Для допустимых значений q случайная величина $T = Y / \sqrt{J}$ имеет функцию распределения $F_T(t)$ следующего вида [5]

$$F_T(t) = \frac{1}{\pi} \int_0^\infty Q_J \left(\frac{1 + \vartheta^2}{2} t^2; q, N \right) \frac{d\vartheta}{1 + \vartheta^2}, \quad t > 0. \quad (6)$$

Для выбранных уровней значимости α правосторонние ($t > 0$) квантили t_α определяются из соотношения

$$\alpha = F_T(\infty) - F_T(t_\alpha) = 1 - \frac{1}{\pi} \int_0^\infty Q_1 \left(\frac{1+g^2}{2} t_\alpha^2; q, N \right) \frac{d g}{1+g^2}. \quad (7)$$

Эти величины отличаются при $q \neq 0$ от стандартных квантилей распределения Стьюдента и переходят в них при $q \rightarrow 0$. Значения уровня значимости α , как правило, достаточно малы, поэтому даже небольшие значения коэффициента корреляции q приводят к заметному изменению $\alpha(t)$ или (при фиксированном уровне α) – к изменению квантиля t_α . Это влияние оказывается достаточно значимым уже при $q \geq 0,2$.

Результаты. На основе (3), (7) создано программное средство, позволяющее построить таблицы квантилей функционала (4) при вариации различных параметров q и N .

В Таблицах 1 и 2 представлены правосторонние квантили с изменяющимся параметром q – коэффициентом корреляции при различных объемах выборки N .

Таблица 1

Правосторонние критические точки t_α функционала Т (4) при $N = 10$, $\sigma = 1$

q	$1 - \alpha$				
	0.900	0.950	0.975	0.990	0.995
0.1	0.880	1.374	1.561	1.815	2.770
0.2	0.882	1.378	1.567	1.823	2.789
0.3	0.885	1.386	1.577	1.837	2.822
0.4	0.890	1.398	1.593	1.859	2.874
0.5	0.898	1.417	1.618	1.892	2.951
0.7	0.928	1.490	1.713	2.023	3.254
0.8	0.961	1.571	1.818	2.166	3.584

С увеличением q график функции распределения имеет тенденцию локализоваться вокруг среднего значения функционала Т, поэтому значения правосторонних квантилей должны уменьшаться, что имеет место в приведенных таблицах.

Анализируя Таблицы 1 и 2, можно также сделать вывод о том, что с увеличением объема выборки N пороговые значения распределения Стьюдента при наличии корреляции уменьшаются. Это означает, что доверительный интервал уменьшается.

Таблица 2

Правосторонние критические точки t_α функционала Т (4), где $N = 20$, $\sigma = 1$

q	$1 - \alpha$				
	0.900	0.950	0.975	0.990	0.995
0.1	0.860	1.326	1.498	1.726	2.531
0.2	0.861	1.329	1.501	1.731	2.542
0.3	0.863	1.333	1.507	1.738	2.560
0.4	0.866	1.340	1.516	1.751	2.590
0.5	0.871	1.350	1.530	1.770	2.635
0.7	0.889	1.394	1.587	1.847	2.818
0.8	0.910	1.445	1.653	1.938	3.029

Далее проанализируем Таблицы 3 и 4.

Таблица 3

Правосторонние критические точки t_α функционала Т (4) при $q = 0.0001$, $\sigma = 1$

N	$1 - \alpha$				
	0.900	0.950	0.975	0.990	0.995
5	0.920	1.476	1.699	2.015	3.365
10	0.879	1.372	1.559	1.813	2.764
15	0.866	1.341	1.517	1.753	2.603
20	0.860	1.325	1.497	1.725	2.528
25	0.856	1.316	1.485	1.708	2.485
30	0.854	1.310	1.477	1.697	2.457

Таблица 4

Правосторонние критические точки t_α функционала Т (4) при $q = 0.6$, $\sigma = 1$

N	$1 - \alpha$				
	0.900	0.950	0.975	0.990	0.995
5	0.949	1.549	1.794	2.145	3.668
10	0.898	1.417	1.618	1.892	2.951
15	0.880	1.373	1.559	1.811	2.739
20	0.871	1.350	1.530	1.770	2.635
25	0.865	1.337	1.512	1.745	2.573
30	0.861	1.328	1.500	1.728	2.532

Из данных этих таблиц, следует, во-первых, что с увеличением уровня значимости α пороговые значения t_α уменьшаются, и, во-вторых, с увеличением объема выборки N квантили также убывают.

С целью верификации результатов рассмотрено распределение Стьюдента [1] в случае отсутствия корреляции (Таблица 5) с числом степеней свободы N в качестве частного случая распределения функционала Т (4). Сравнивая строки в Таблицах 3 и 5 при $N=10$ можно убедиться в правильности построения таблиц.

Таблица 5

Табличные правосторонние критические точки t_α функционала Стьюдента

N	$1 - \alpha$				
	0.900	0.950	0.975	0.990	0.995
5	0.920	1.476	1.699	2.015	3.365
10	0.879	1.372	1.559	1.813	2.764
15	0.866	1.341	1.517	1.753	2.602
20	0.860	1.325	1.497	1.725	2.528
25	0.856	1.316	1.485	1.708	2.485
30	0.854	1.310	1.477	1.697	2.457

На рис. 1-2 приведены плотности и функции распределения для функционала (4) соответственно.

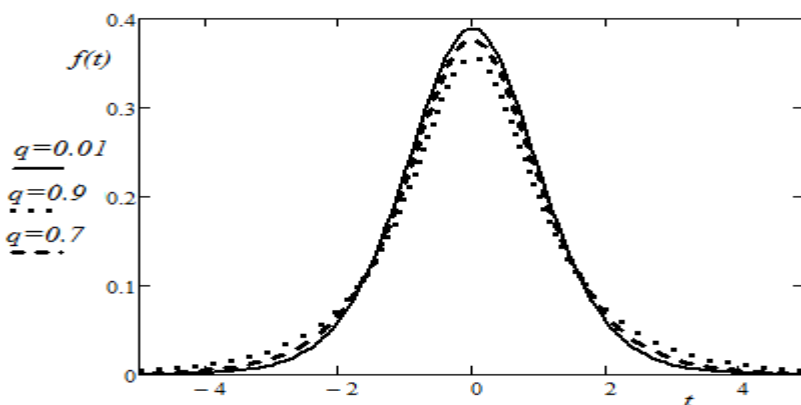


Рис. 1. Плотности распределения вероятностей для функционала (4) с параметрами: $q = 0.01, 0.7, 0.9, N = 10, \sigma = 1$

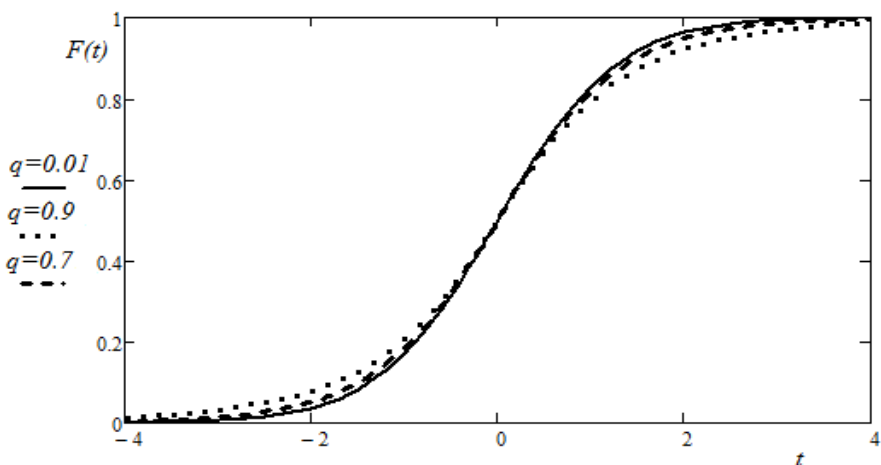


Рис. 2. Графики функции распределения вероятностей для функционала (4) с параметрами: $q = 0.01, 0.7, 0.9$, $N = 10$, $\sigma = 1$

Выводы. В работе впервые получены значения квантилей распределения Стьюдента для случая коррелированных отсчетов. Приведенные таблицы правосторонних пороговых значений t_α могут использоваться для контроля. Полученные результаты подтверждают адекватность модели и могут быть использованы при анализе опытных данных.

Найденные выражения для распределений стандартизованного отношения Стьюдента допускают предельный переход $N \rightarrow \infty$ для отсчетов случайного процесса на заданном интервале. Возможно также обобщение результатов на случаи комплекснозначных или многомерных случайных переменных.

Список литературы: 1. *Ивченко Г. И., Медведев Ю. И.* Математическая статистика. – М.: Высшая школа, 1984. – 248 с. 2. *Пугачев В. С.* Теория вероятностей и математическая статистика. М.: Наука, 1979. – 496 с. 3. *Рытов С. М.* Введение в статистическую радиофизику. Харьков: Высшая школа, 1990. – С. 5–7. 4. *Тихонов В. И., Миронов М. А.* Марковские процессы. – Вестник НТУ «ХПИ». Сб. научных трудов. Тем. вып. «Системный анализ, управление и информационные технологии». – Харьков: НТУ «ХПИ». – 2005. – № 41. – С. 220–227. 5. *Мазманишвили А. С.* Коррелированные последовательности в задачах статистического оценивания // Доповіді Національної Академії Наук України, 2001, № 7, с.21–25.

Поступила в редколлегию 11.02.09